Digital Signal Processing 133 (2023) 103836

Contents lists available at ScienceDirect

Digital Signal Processing

journal homepage: www.elsevier.com/locate/dsp

Long-memory parameter estimation based on fractional spline wavelets

Mateus Gonzalez de Freitas Pinto *, Chang Chiann

Institute of Mathematics and Statistics, University of São Paulo, Rua do Matão, 1010, São Paulo, 05508-090, São Paulo, Brazil

ARTICLE INFO

Article history:
Available online 21 November 2022

Keywords:
Long-memory
Hurst exponent
Wavelets
Splines
Fractional splines
Fractional random walk

ABSTRACT

Fractional splines extend Schoenberg's B-splines to fractional orders, which were proven to fulfill all requirements to form wavelet bases. Moreover, some of these fractional splines wavelets act approximately as fractional difference operators for signals that are essentially lowpass and concentrated around the origin, which makes them helpful in analyzing series with fractal behavior. In this article, we propose a novel estimator for the long memory parameter of a time series based on the fractional spline discrete wavelet transform (FrDWT), using the fact that it works approximately as a fractional differentiator. We perform simulations and examples to illustrate the proposed method. The proposed estimator outperforms traditional and widely used estimators in simulated data. We also show that, as the sample size increases, the bias in the simulation tends to decrease, as well as the variance, which might indicate that the estimator is consistent.

1. Introduction

Long memory models were developed in the 1950's in studies for climatology and hydrology, only then being later extended to other fields of knowledge, such as economics, astronomy, chemistry, finance, geosciences and statistics. This phenomenon was noted by [1], and the first model for long range dependence was proposed in [2], with fractional Brownian motion and fractional noise, with heuristics to estimate the self-similarity parameter in [3]. Applications were then identified for the self-similar process to economics in [4], where self-similar processes, at a sufficiently good approximation, had a spectral density typically with a pole at the origin. This led to the later development of the ARFIMA (AutoRegressive Fractionally Integrated Moving Average) models [5,6].

A long memory process is characterized by an autocorrelation function (ACF) that decays hyperbolically towards zero, with such a slow decay that their autocorrelations are not summable [7]. An alternative way to characterize this process is using its spectral density as in [8], with a power spectrum having a pole at zero frequency. Alternatively, [5] models long-range dependent processes using the fractional integration parameter operator.

Numerous data sets exhibit long-range dependence, noted by the slow decay of their ACF or by the inspection of their power spectral densities. Some examples are the temperature data in [9] and hydrology-related time series in [8]. For other examples in different fields of study, see [10,11].

Due to this wide range of applications, several estimators have been developed for the d order of the long memory or the Hurst parameter H, which differs from d up to a constant. In the scope of this article, we discuss estimators for the fractional random walk (FRW) and the fractional white noise (FWN), or pure long memory process.

One of the oldest and best-known methods is the R/S statistics for the Hurst exponent [1], sometimes referred to as the rescaled range. Although very popular in usage, the R/S statistics is known for producing biased estimates for the Hurst exponent, especially when the sample size is small. Therefore, there is also the corrected R/S estimator of [12], as a slight modification of the proposition of [13], named R/S-AL estimator.

There are other approaches to estimate the fractional difference parameter. For instance, a spectral approach of [14] uses the smooth periodogram. There is also the Geweke-Porter-Hudak (GPH) estimator, which is a semi-parametric approach. It is an ordinary least square estimator based on the simple linear regression of the log periodogram on a deterministic regressor [15]. Moreover, the Whittle approximate MLE (Maximum Likelihood Estimation) is also a periodogram-based method that assumes standard normality for the disturbance [16,8].

Although wavelets have been used to estimate the long memory parameter [17,18], there are not yet, to the best of our knowledge, estimators based on fractional spline wavelets for the long memory parameter or the Hurst exponent.

* Corresponding author.
E-mail address: mateusgf@ime.usp.br (M. Gonzalez de Freitas Pinto).

https://doi.org/10.1016/j.dsp.2022.103836
1051-2004/

Fractional spline wavelets were introduced in [19,20]. The family of fractional splines of order α involves the linear combination between the one-sided power function and the fractional difference operator, and have a fractional order of approximation of $\alpha+1$ while reproducing polynomials of degree $\lceil\alpha\rceil$, extending the proposition of B-splines in [21] for fractional orders.

Moreover, if $\alpha > -1/2$ to fulfill square integrability, they satisfy the Riesz bounds and the two-scale relations, which are the requirements for multiresolution analysis of L_2 [19,22]. Another attractive property is that fractional spline wavelets behave like fractional differentiators, so they help analyze $1/f^{\beta}$ noises [20] and signals that have a spectral density essentially low-pass concentrated around the origin, since by an appropriate choice of both α and weights in the wavelet channel, the $1/f^{\beta}$ signal can be whitened [23].

We propose an estimator based on the fractional spline discrete wavelet transform (FrDWT) to the d parameter for the Fractional Gaussian Noise (FGN) and for the Fractional Random Walk (FRW), using the property identified in [20] that these fractional spline wavelets behave essentially as fractional differentiators and the fact that with a suitable weighting in the wavelet channel for white noise, we can synthesize a $1/f$-like noise [23]. We conversely use these properties to whiten the input FRW signal, using a relation between the order of the fractional spline wavelet and the long memory parameter of the input FRW. Following the procedure further detailed in the methodology section, we choose the order that best whitens the input signal as an estimate for the long memory parameter.

We perform simulation studies which show the behavior of the FrDWT-based estimator, which tends to be less biased and have a smaller variance than other estimators. Moreover, the proposed method has smaller MSE when compared to other estimation procedures. Lastly, we perform practical applications in three real signals with long-range dependence and show that the proposed procedure performs best in terms of Root Mean Squared Error (RMSE), Mean Absolute Percent Error (MAPE) and Mean Absolute Error (MAE). Assessing the results of both simulations and empirical applications, we find the performance of the FrDWT-based estimator to be satisfactory.

This paper is divided into the following sections: in the second one, we discuss some backgrounds on wavelets, splines, and fractional spline wavelets, which will be used to construct the estimator. In the third section, we discuss the methodology and give some examples of the proposal. The fourth section consists of simulations to assess the estimator empirically, as well a part of empirical applications with real time series that exhibit the pattern of long-range dependence. Finally, there is a section of final remarks and further work discussion.

2. Backgrounds

2.1. Wavelets, splines and spline-wavelets

Wavelets are functions $\psi \in L_2$ with $\int_{\mathbb{R}} \psi(t)dt = 0$, $||\psi|| = 1$ and centered in the neighborhood of $t = 0$. Wavelets are localized in time and scale, making them ideal for analyzing signals with fractal structures. It is possible to construct wavelets such that the dilated and translated families form an orthonormal basis of L_2. The wavelet basis are generated from binary re-scales and translations the ψ mother wavelet [24].

For a mother and father wavelets, we set:

$$\psi_{j,k}(t) = 2^{j/2}\psi(2^j t - k), \tag{1}$$

$$\phi_{j,k}(t) = 2^{j/2}\phi(2^j t - k), \tag{2}$$

where $j, k \in \mathbb{Z}$ [25].

A sequence of nested, closed subspaces V_j of L_2, for $j \in \mathbb{Z}$, forms a multiresolution analysis in the sense of [22] if

1. $\ldots \subset V_{-1} \subset V_0 \subset V_1 \subset \ldots$;
2. $L_2 = \overline{\bigcup_{j=-\infty}^{\infty} V_j}$;
3. $\bigcap_{j=-\infty}^{\infty} V_j = \{0\}$;
4. $f(x) \in V_j \Leftrightarrow f(2x) \in V_{j+1}$;
5. There exists $\phi \in V_0$ such that $\{\phi(x-k)\}_{k\in\mathbb{Z}}$ forms a Riesz basis of V_0.

If the mother and father wavelet families fulfill the requirements for a multiresolution analysis, using equations (1) and (2), for any fixed j_0, a multiresolution decomposition in $L_2(\mathbb{R})$ can be written as

$$L_2(\mathbb{R}) = \bigoplus_{j\in\mathbb{Z}} W_j = V_{j_0} \oplus \bigoplus_{j\geq j_0} W_j, \tag{3}$$

where W_j is the subspace generated by $\{\psi_{j,k}\}_{k\in\mathbb{Z}}$ and V_{j_0} is the subspace generated by $\{\phi_{j_0,k}\}_{k\in\mathbb{Z}}$. Any discrete signal or time series $\{X_t\}_{t=1,2\ldots T}$ with $T = 2^J$ for J integer, can be represented as

$$X_t = c_{0,0}\phi(t) + \sum_{j=0}^{J-1}\sum_{k=0}^{2^j-1} d_{j,k}\psi_{j,k}(t), \tag{4}$$

with $j_0 = 0$ and coefficients given by the inner product $d_{j,k} = \langle X, \psi_{j,k}\rangle$ and $c_{j,k} = \langle X, \phi_{j,k}\rangle$.

Splines were introduced originally in [21], but gained significant attention only after 1960 with applications in several fields of study in physics and applied mathematics [26–29]. They are piecewise polynomials within specific intervals, with knots connecting them and high-order continuity at the knots. A polynomial spline of degree n needs $n+1$ coefficients to describe each of the pieces [30].

An interesting link between splines and wavelet theory is discussed in [31]. For instance, polynomial splines constitute the only wavelets that have the explicit analytical form [20]. Therefore many of the earlier wavelets were based on splines [30]. Moreover, B-splines follow some of the fundamental wavelet properties, such as: order of approximation, reproduction of polynomials, vanishing moments, multiscale differentiation property, and smoothness of the basis functions, which makes them suitable for this purpose [31].

The spline wavelets approach was proposed by [32], showing that spline spaces generated a multiresolution analysis, in the sense of [22], and constructing wavelet bases following the traditional techniques, having as a particular case the Haar wavelets. For a more detailed review on splines see [33,30] and their relation with wavelets [34,31].

2.2. Fractional spline wavelets

This article uses the fractional spline wavelets developed in [19]. They result in an extension of the family of polynomial splines of [21] for non-integer degrees $\alpha > -1/2$, representing the Hölder exponent of the fractional B-spline. It is possible to show that these fractional splines can be used as basis functions and that for $\alpha > -1/2$ they form a Riesz basis and satisfy the requirements for multiresolution analysis in the sense of [22] for the L_2 Hilbert space [19,20].

Wavelets help to detect and analyze singularities in signals because wavelets with n vanishing moments behave as $(n+1)$-order differentiator: $\hat{\Psi}(\omega) \propto \omega^{n+1}$, as $\omega \to 0$. However, when dealing with long memory models, such as a fractional white noise or a fractional random walk, it is better for a wavelet to behave as

a fractional differentiator: $\hat{\Psi}(\omega) \propto \omega^{\alpha+1}$, where α is non-integer [23].

The functions in [19] to be used to form a basis are the fractional B-splines, which are localized versions of the one-sided power function: $(x-k)_+^\alpha := \max\{x-k, 0\}^\alpha$. The causal fractional B-spline is defined as

$$\beta_+^\alpha(x) := \frac{\Delta_+^{\alpha+1} x_+^\alpha}{\Gamma(\alpha+1)} = \sum_{k\geq 0} \frac{(-1)^k \binom{\alpha+1}{k}}{\Gamma(\alpha+1)} (x-k)_+^\alpha, \tag{5}$$

where $\alpha > -1/2$ in order for square integrability to hold and $\Delta_+^{\alpha+1}$ is the fractional forward finite difference operator

$$\Delta_+^\alpha f(x) := \sum_{k\geq 0} (-1)^k \binom{\alpha}{k} f(x-k). \tag{6}$$

The operator in (6) is the discretization of Liouville's generalization of the differentiation of fractional orders [35], and is the same operator used when dealing with long memory processes, such as the ARFIMA or the Fractional Random Walk [7]. Anti-causal and symmetrical fractional B-splines are also defined in [19].

It is shown in [19] that the frequency response of the scaling filter H^α is given by

$$H_+^\alpha(e^{j\omega}) = \sqrt{2}\left(\frac{1+e^{-j\omega}}{2}\right)^{\alpha+1}, \tag{7}$$

for the causal fractional B-splines. The standard notation from now on for causal fractional order B-splines will be respectively with '+'.

The discrete fractional spline wavelet transform is calculated using the FFT-based algorithm of [23]. There, the authors generate the orthonormal wavelet filters using the technique in [36]:

$$H_\perp^\alpha(e^{j\omega}) = H^\alpha(e^{j\omega}) \sqrt{\frac{A^{2\alpha+1}(e^{j\omega})}{A^{2\alpha+1}(e^{2j\omega})}}, \tag{8}$$

where $A^\alpha(z)$ is the autocorrelation filter of α fractional B-spline:

$$A^\alpha(e^{j\omega}) = \sum_{n\in\mathbb{Z}} e^{-jn\omega} \int \beta_+^\alpha(x)\beta_+^\alpha(x+n)dx. \tag{9}$$

For the computation of (9) for an approximation of the integral with a sum of N terms, it is used an asymptotic Poisson-equivalent formula of [23] that we use in the implementation to compute the filters.

From this, using classical wavelet theory, it is possible to obtain the frequency response of the generating filter from the orthogonal case by [24]:

$$G_\perp^\alpha(e^{j\omega}) = e^{-j\omega} H_\perp^\alpha(-e^{-j\omega}), \tag{10}$$

from which one constructs fractional spline wavelets by the linear combination of the scaling functions [23].

Technicalities about the construction of such wavelet bases and multiresolution analysis using them can be seen in [19,20], whereas in [23] one can find more details about the algorithm for wavelet analysis and synthesis using fractional spline wavelets and implementation of the discrete fractional spline wavelet transform. Other technicalities concerning splines can be seen in [34].

3. Methodology

An interesting and remarkable feature is that, as stated in [20], the fractional spline wavelet behaves like a fractional derivative operator because $\hat{\Psi}_+^\alpha(\omega) = C(j\omega)^{\alpha+1}, \omega \to 0$ and $\hat{\Psi}_*^\alpha(\omega) = C|\omega|^{\alpha+1}, \omega \to 0$. In this case, the wavelet coefficients can be understood as the samples of the $(\alpha+1)$-th fractional derivative of a series of reduced resolution versions of the input signal. Therefore, the transform acts as a fractional differentiator for signals X_t with a spectrum $S_X(\omega)$ low-pass concentrated around the origin. This is precisely the case for signals as the FWN and FRW [7], which will be a core concept for the proposed estimator.

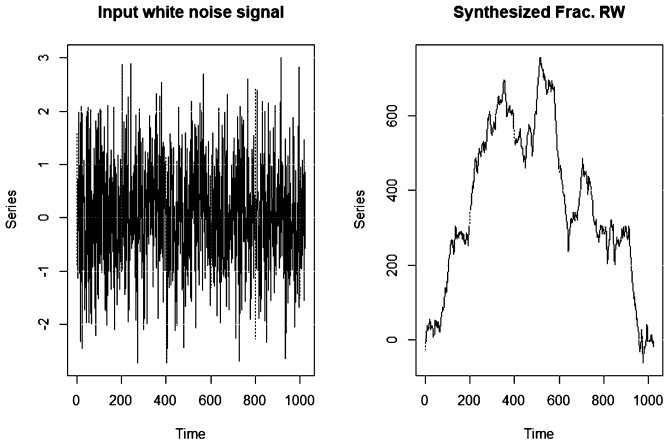


Fig. 1. Synthesis procedure example.

In [23], the authors provide a way to synthesize a 1/f noise by using as an input a white noise through the wavelet channels and weight them by $2^{j(\alpha+1)}$ in the $j-$th band. Indeed, this proposal makes sense for the case of long memory processes, because in [17] it is shown that for an integrated process $I(d)$ with $|d| < 1/2$, the wavelet decomposition coefficients, denoted $d_{j,k}$, are such that

$$d_{j,k} \sim \mathcal{N}(0, \sigma^2 2^{-2jd}), \tag{11}$$

where σ^2 is finite and depends on the integral $\int \psi(s)\psi(s-t)ds$ of a given mother wavelet.

Consider the following example, adapted from [23] for illustrative purposes. Here, we synthesize a fractional random walk signal, using white noise in the wavelet channels as an input. We consider $d = 0.25$. Thus we look for a signal that follows $\Delta^{1+d} X_t = \varepsilon_t$. To generate this, we weight the noise by $2^{j(\alpha+1)}$ in the j-th level of the decomposition.

First, we simulate a white noise with $T = 2^{10}$ observations. Then, we use the fractional spline DWT (FrDWT) and weight the coefficients by $2^{j(\alpha+1)}$ in the j-th level of the decomposition for each decomposition level. We then reconstruct the signal, which should be a trajectory of a fractional random walk with parameter $1+d$.

In Fig. 1, we can see the input signal, a white noise (left) that behaves essentially as a fractional random walk after the weighting procedure in the wavelet channel (right). This can be confirmed by seeing their autocorrelations in Figs. 2 and 3, because while the input signal has typical sample ACF and partial autocorrelation function (PACF) of a white noise, both the ACF and PACF of the output signal exhibit the typical decay of a fractional random walk. The same happens for the spectrum in Fig. 4: while the spectrum of the input series follows the pattern expected for a white noise, the synthesized series displays the typical behavior of a time series with long memory. Indeed, when using the GPH estimator to determine the long-memory parameter, the estimated value is $\hat{d} + 1 = 1.222$.

Since we look for an estimator of the long memory parameter d, we consider the reverse problem of the example, meaning: having an input of a fractional random walk, we search for suitable fractional-order α^*, and weights $2^{j\beta(\alpha^*)}$ applied to the wavelet decomposition levels such that the wavelet reconstruction with same

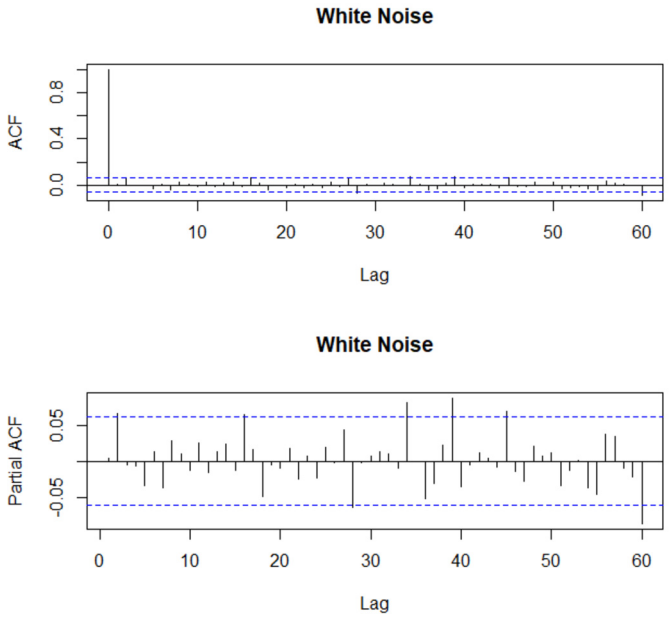


Fig. 2. Sample ACF and PACF for the white noise.

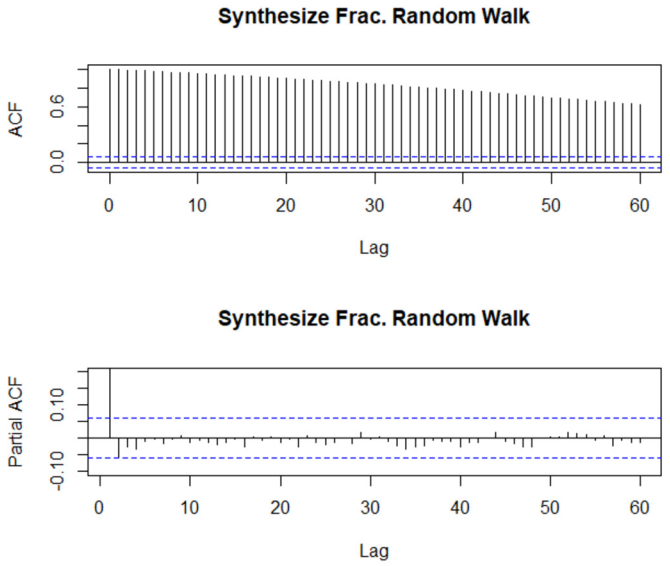


Fig. 3. Sample ACF and PACF for the synthesized series.

fractional-order α^* returns a "whitened" signal, where $\beta(\alpha)$ is a function of α. The heuristics of the choice of relying on the properties of the FrDWT.

Thus consider $\{X_t\}$ with $t = 1, 2, \ldots T$ and $T = 2^J$, $J \in \mathbb{Z}_+$ a signal that is generated according to a fractional random walk,

$$\Delta_+^{d+1} X_t = \varepsilon_t, \tag{12}$$

where $\varepsilon_t \sim N(0, 1)$ with parameter d where $|d| < 1/2$. Assume that the parameter d is known and let $d_{j,k}$ be the coefficients associated with the wavelet decomposition of order α

$$d_{j,k} = \langle \psi_{j,k}^{\alpha}, X_t \rangle. \tag{13}$$

A question that arises is the order α of the fractional spline wavelet to be used in the procedure that whitens this input signal. Since the fractional spline wavelet transform works essentially as a fractional differentiator of order $\alpha + 1$ [20], let $\alpha^* = \frac{(d+1)}{2}$ be the

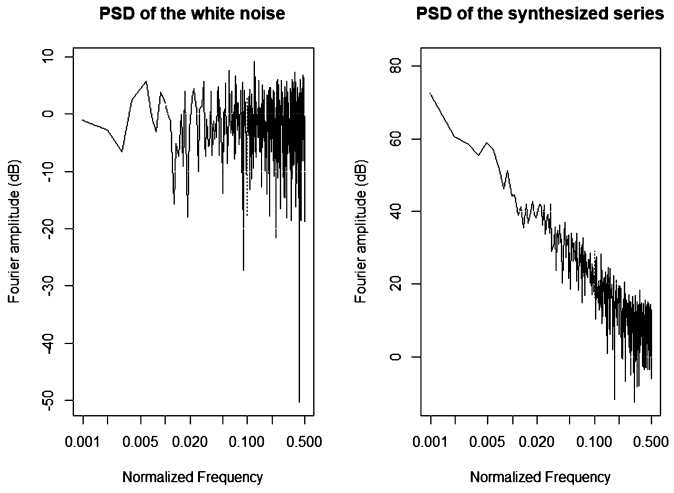


Fig. 4. Power spectral density (PSD) for the synthesized series and white noise.

fractional order of the wavelet. Then $\psi_{j,k}^{\alpha^*}$ behaves like a differentiator of order $\left(\frac{d}{2} + \frac{3}{2}\right)$. Because $X_t \sim I(d+1)$, then after applying the FrDWT, the integration order should be $\frac{d}{2} - \frac{1}{2}$. This has associated coefficients

$$d_{j,k} \sim \mathcal{N}\left(0, \sigma^2 2^{-2j\left(\frac{d}{2} - \frac{1}{2}\right)}\right), \tag{14}$$

and after weighting each level of the decomposition with $2^{-j(d+1)}$, it leads to

$$2^{-j(d+1)} d_{j,k} \sim \mathcal{N}\left(0, \sigma^2 2^{-2j\left(-\frac{d}{2} - \frac{3}{2}\right)}\right). \tag{15}$$

The equation (15), on the other hand, corresponds to the distribution of the coefficients of a process with integration order $I\left(-\frac{d}{2} - \frac{3}{2}\right)$. Now because the fractional reconstruction operates as an integrator operator of order $\left(\frac{d}{2} + \frac{3}{2}\right)$, then the resulting reconstructed signal should be $I(0)$, which is white noise.

This heuristic is analogous for the fractional Gaussian white noise. Let X_t be a fractional white noise, meaning $\Delta^d X_t = \varepsilon_t$ for a known d and let $d_{j,k}$ as in (13). If we choose $\alpha^* = \frac{d}{2}$, then the fractional spline wavelet transform behaves as a fractional differentiator of order $\frac{d}{2} + 1$. Now since $X_t \sim I(d)$, then the resulting order of integration after applying the FrDWT should be $\left(\frac{d}{2} - 1\right)$, which has associated coefficients

$$d_{j,k} \sim \mathcal{N}\left(0, \sigma^2 2^{-2j\left(\frac{d}{2} - 1\right)}\right), \tag{16}$$

and after weighting each level of the decomposition with 2^{-jd}

$$2^{-jd} d_{j,k} \sim \mathcal{N}\left(0, \sigma^2 2^{-2j\left(-1 - \frac{d}{2}\right)}\right), \tag{17}$$

that corresponds to the distribution of the DWT coefficients of integration order $I\left(-\frac{d}{2} - 1\right)$. Now, as with the FRW, since the fractional reconstruction for order $\alpha^* = d/2$ operates as fractional integration operator of order $\frac{d}{2} + 1$, the resulting reconstructed signal should be $I(0)$, which is a white noise.

This whitening procedure described above can be described by the Algorithm 1. Note that the only difference between whitening a FRW and a FGN is the exponent of 2^x, either $-j(\alpha + 1)$ or $-j\alpha$ that is applied to the wavelet coefficients $d_{j,k}$.

Algorithm 1 Procedure to whiten a FRW or FGN.

Input: A time series $\{X_t\}$ for $t = 1, 2, \dots T$. A fractional order α
Output: A whitened time series $\{W_t^{\alpha}\}$ for $t = 1, 2, \dots T$.

```
procedure WhitenSignal(X_t, α)
    Decompose X_t to the wavelet coefficients d_{j,k}, j = 0, 1, ... J − 1 with d_{j,k} = ⟨X, ψ_{j,k}^{(α+1)/2}⟩.
    for j = 0, 1, ... J − 1 do
        if input signal is FRW then
            Compute d̃_{j,k} = 2^{−j(1+α)} d_{j,k}
        else if input signal is FGN then
            Compute d̃_{j,k} = 2^{−jα} d_{j,k}
        end if
    end for
    Reconstruct W_t^α, the signal with coefficients d̃_{j,k}.
    return W_t^α
end procedure
```

Let us assume that the parameter $d \in (-0.5, 0.5)$ is unknown, and we have the interest to estimate it. Consider that the observed series has a data generating process of a FRW$(1+d)$ with an underlying Gaussian distribution. Then we can search for an estimator to the fractional order d as being the value $\hat{d}$ such that the output signal is "best whitened" after following Algorithm 1 with an order $\hat{d}$ to the input signal.

To formalize our argument, consider an input signal $\{X_t\}, t = 1, 2, \dots 2^J$ and $X_t \sim \text{FRW}(1+d)$, and assume d is unknown. Define the thresholding function that applies the weights for every $d_{j,k}$ coefficient as

$$\eta_j(x) = x2^{-j(\alpha+1)}, \text{ for } j = 0, 1, \dots J-1. \tag{18}$$

The goal is to estimate the unknown d parameter.

Consider the search across the values $\alpha \in (-0.5, 0.5)$. We decompose the input signal for the order $(\alpha+1)/2$ of the fractional spline wavelet decomposition. We perform the weighting procedure in (18) to the coefficients, $\eta_j(d_{j,k}) = 2^{-j(\alpha+1)} d_{j,k}$ for $j = 0, 1, \dots J-1$. With this, reconstruct the signal using the inverse fractional spline wavelet transform with the same $(\alpha+1)/2$ order of the fractional splines wavelet decomposition. This is the whitening procedure described in Algorithm 1.

Let W_t^{α} denote the output signal of the whitening procedure of Algorithm 1. A way to measure the adequacy of α as an estimator for d is by assessing the level of "whiteness" of W_t^{α}. We do so by considering the sum of the n first squared sample autocorrelations, meaning:

$$R_n^{\alpha} = \sum_{k=1}^{n} \hat{\rho}_w^2(k), \tag{19}$$

where $\hat{\rho}_w(k) = \hat{\gamma}_w(k)/\hat{\gamma}_w(0)$ and

$$\hat{\gamma_w}(k) = \frac{1}{T} \sum_{t=1}^{T-|k|} (W_{t+|k|}^{\alpha} - \overline{W}^{\alpha})(W_t^{\alpha} - \overline{W}^{\alpha}). \tag{20}$$

Therefore, an estimator could be obtained as

$$\hat{d} = \arg\min_{\alpha} R_n^{\alpha}. \tag{21}$$

The objective function in (21) is reasonable because, $R_n^{\alpha} = \sum_{k=1}^{n} \hat{\rho}_w^2(k) \geq 0$. In fact, if $\alpha = d$, then $W_t^{\alpha+1} \approx \Delta^{1+d} X_t = \varepsilon_t$ by the definition of the FBM. Therefore $\rho_w(k) = \rho_{\varepsilon}(k) = 0, k \neq 0$ and thus as $\alpha \to d$, we have $R_n^{\alpha} \to 0$, which is the minimum value for R_n^{α}.

On the other hand, assume now that $\alpha \neq d$, which leads to, $W_t^{1+\alpha} \approx \Delta^{1+\alpha} X_t$. But since $X_t = \Delta^{-d-1}\varepsilon_t$, $W_t^{1+\alpha} \approx \Delta^{1+\alpha-d-1}\varepsilon_t = \Delta^{\alpha-d}\varepsilon_t$, which is a FGN. Therefore $\rho_w^2(k) := \gamma_w^2(k)/\gamma^2(0) > 0$, because $\gamma_w^2(k) > 0$ for the FGN. Thus $R_n^{\alpha} = \sum_{k=1}^{n} \hat{\rho}_w^2(k) \geq 0$ and will not be null except when $\alpha = d$, which explains the choice of the objective function in (21).

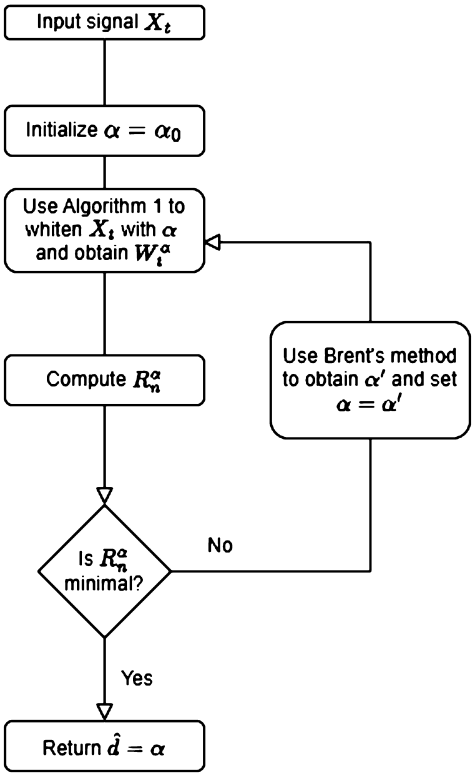


Fig. 5. Box diagram describing the optimization procedure.

To search the optimal α for (21), we must use a numerical optimization algorithm. We find Brent's algorithm for one-dimensional section search and successive parabolic interpolation suitable for our purposes [37]. The estimation procedure can be summarized in the box diagram of Fig. 5.

A question that arises from the procedure is the choice of n. We consider n as a function of sample size $T = 2^J$, but still as a constant for every $J \in \mathbb{Z}_+$. There is no consensus in the literature about the size of n, although there are recommendations, such as [38–42]. Specifically in [42] a simulation study was performed for some sample sizes and measured, under specific conditions, a suitable value for n when dealing with Ljung-Box tests.

Because of that, for our simulations with $J = 10$, we fix $n = 25$. This relies on the fact that in [38] it is suggested to consider $n = 20$, and simulation studies performed in reference [42] have shown that for series with length $T = 1,000$ and significance level 0.01, an appropriate choice is $n = 25$. The choice for values of n when $J \neq 10$ will be explained in due course.

4. Simulation studies

We performed simulations comparing this FrDWT-based estimator with those described in the introduction. For our purposes, we denote the GPH estimator [15] as "d.GPH"; the periodogram-based estimator [14] as "d.Sperio"; the Whittle estimator [8] as "d.Whittle"; the rescaled range (R/S) [1] as "RoverS"; the corrected AL-R/S estimator [13] as "CorrectRS"; and finally the proposed estimator using the fractional spline wavelet transform, denoted as "d.FrWT". The wavelet family chosen was generated by causal orthogonal fractional splines because $\hat{\Psi}_+^{\alpha}(\omega) = C(j\omega)^{\alpha+1}, \omega \to 0$. All the computation was performed using R [43].

We simulate 1000 series of a fractional random walk, meaning $\Delta^{1+d} X_t = \varepsilon_t$, $\varepsilon_t \sim \mathcal{N}(0, 1)$ for $t = 1, 2, \dots 1024$ with parameters d

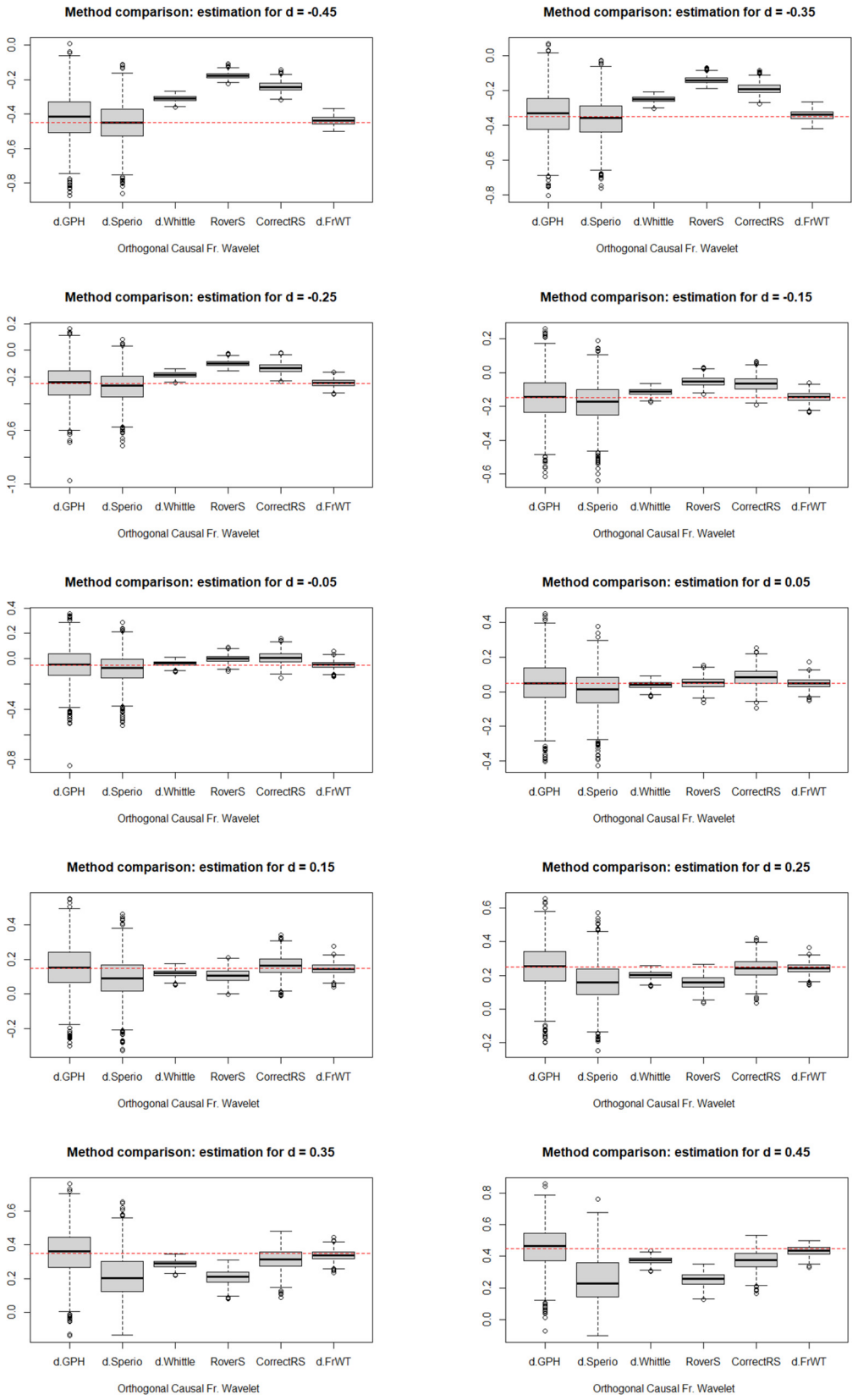


Fig. 6. Boxplot of estimated d with $J = 10$ for different values of d.

taking values in $\{-0.45, -0.35, -0.25, -0.15, -0.05, 0.05, 0.15, 0.25, 0.35, 0.45\}$. The goal is to estimate the long memory parameter d using the proposed methodology and compare it with the other estimators to assess its performance. For means of comparison, we consider the series with length $T = 2^{10}$, but we will also show the behavior of the proposed estimator for different series with length $T = 2^J$, for $J = 7, 8, 9, 10, 11$.

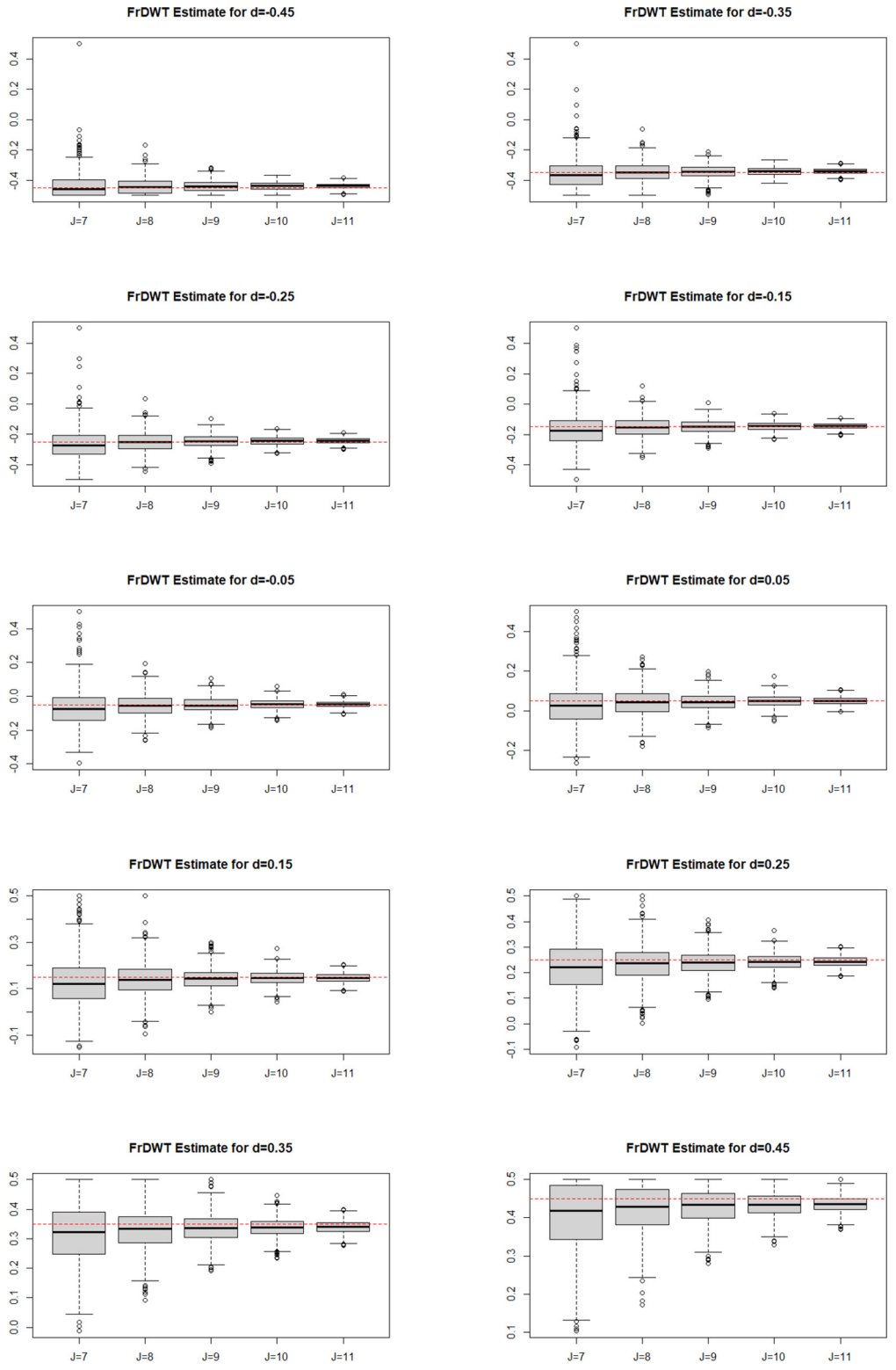
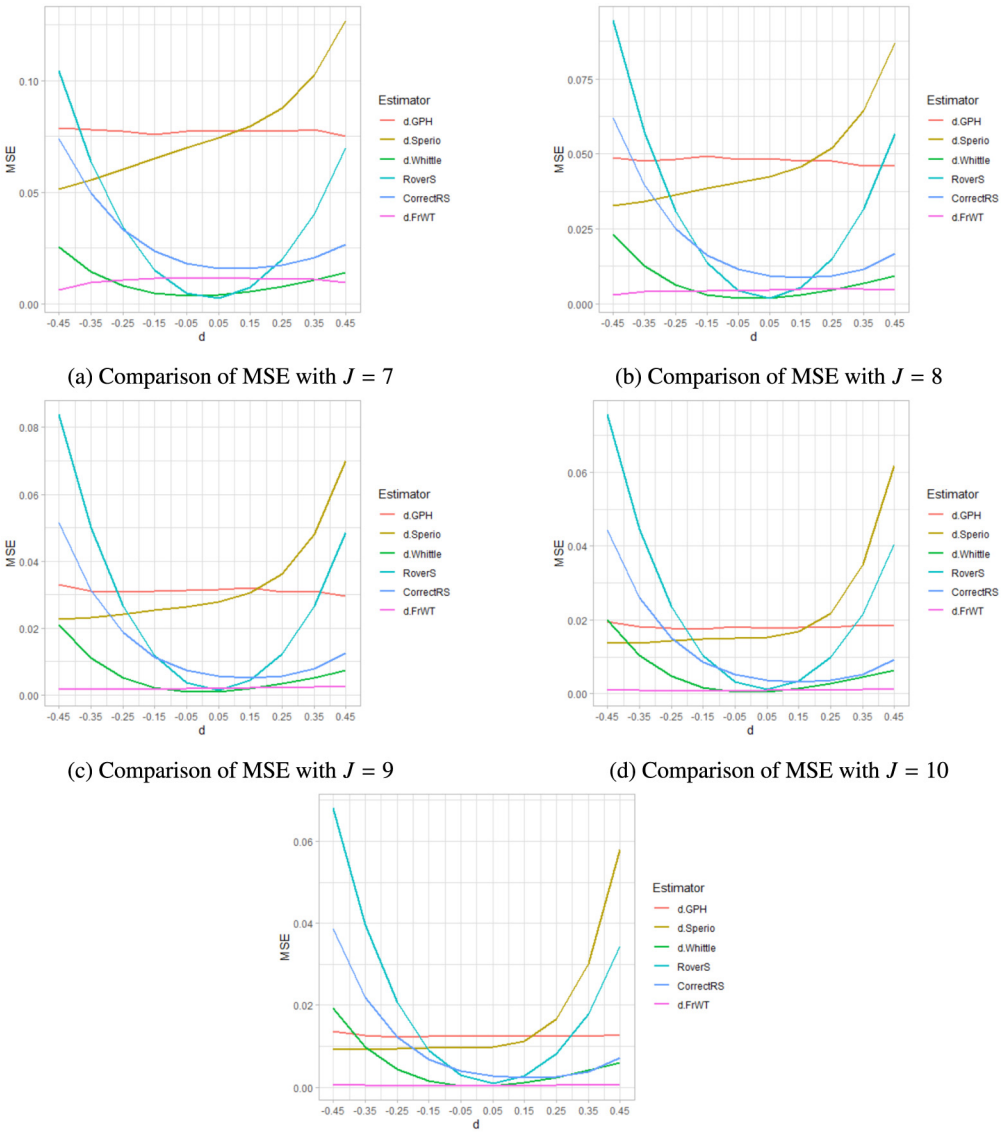


Fig. 7. Boxplot of estimated d with increasing J for different values of d.

It is worth mentioning that we do not simulate the fractional random walk using the synthesis procedure exemplified in the methodology section. The fractional random walk data were simulated following [44,45].

Figs. 6 display the boxplot of the estimated values of d, comparing the different approaches. It is clear from the simulations that the estimator is less biased than the others for all values of d. Moreover, the FrDWT-based estimator has a much smaller variance compared with the less biased approaches (the GPH and periodogram-based estimator).

The FrDWT-based estimator for d outperforms the others in the simulation study. Furthermore, although the Whittle estimator's values of $|d| = 0.05$ seem competitive, as $|d|$ increases, the Whittle estimator becomes biased, whereas the proposed FrDWT estimator remains almost unbiased. Therefore, we perceive that the proposal is more suitable for general and practical usage.



(a) Comparison of MSE with $J = 7$

(b) Comparison of MSE with $J = 8$

(c) Comparison of MSE with $J = 9$

(d) Comparison of MSE with $J = 10$

(e) Comparison of MSE with $J = 11$

Fig. 8. Comparison of MSE between estimators when varying J.

We also consider the estimates for different J values, meaning $J = 7, 8, 9, 10, 11$. The heuristics for determining the value of n in (19) is the same, based on the literature recommendations and in the simulation studies of [42]. We set $n = 15$ and $n = 20$ for $J = 7$ and $J = 8$, respectively. For $J = 9, 10, 11$, $n = 25$ is fixed.

Table 1 displays the mean bias and the mean squared errors (MSE) of the estimates for d for the different values of J. We calculate the MSE as

$$\text{MSE} = \frac{1}{N}\sum_{k=1}^{N}(\hat{d}_k - d)^2, \tag{22}$$

where N is the number of simulated series, $\hat{d}_k$ is the estimated value of d for the $k-$th sample and d is the true value used to generate the simulated series. In our case, $N = 1000$.

Fig. 7 shows that, in the simulated data, the estimator tends to be less biased as J increases and has a decreasing variance, which might indicate the consistency of the estimator. It is worth mentioning that even for small sample sizes such as $T = 128$ ($J = 7$), the estimator has a mean bias that is significantly close to 0 for all values of d. It is noteworthy that, for all J, the bias tends to increase as $|d| \to 0.5$, mainly because it is approaching the unstable non-stationary behavior that defines these processes for such values of $|d| > 0.5$. Overall, the bias tends to be close to 0, and the standard deviation decreases with J increasing for all values of d.

We also compare the MSE when estimating d with different procedures and increasing the value of J, for $J = 7, 8, 9, 10, 11$. This is illustrated in Fig. 8, where we plot the MSE for each value of d and for each estimator. It is possible to see that the FrDWT-based estimator's performance increases with J, meaning that for bigger sample sizes such as $T = 2048$, the MSE is very close to 0 for all values of d, outperforming the estimators we compared it with, as it can be seen in Figs. 8c, 8d and 8e. Even for smaller samples ($T = 128$ and $T = 256$), the only competitive estimators are the Whittle estimator for $|d| \leq 0.25$ and the R/S statistics when $|d| = 0.05$, as it can be seen in Figs. 8a and 8b.

Therefore, the FrDWT-based estimator seems like a reasonable and competitive approach to estimate the long memory parameter. It is worth noting that the proposed estimator performs well for both persistent and anti-persistent long memory patterns in the time series, which is also desirable, even though anti-persistent series are less commonly seen in practice.

Table 1
Comparison of mean bias and MSE for the proposed method as J increases.

J	Mean Bias $d=-0.45$	MSE $d=-0.45$	Mean Bias $d=-0.35$	MSE $d=-0.35$
7	0.01313	0.00618	−0.00900	0.00950
8	0.01214	0.00285	0.00220	0.00392
9	0.01125	0.00159	0.00683	0.00176
10	0.01104	0.00091	0.00876	0.00090
11	0.01166	0.00049	0.00986	0.00045

J	Mean Bias $d=-0.25$	MSE $d=-0.25$	Mean Bias $d=-0.15$	MSE $d=-0.15$
7	−0.01599	0.01072	−0.01873	0.01137
8	−0.00079	0.00409	−0.00337	0.00424
9	0.00411	0.00177	0.00139	0.00179
10	0.00665	0.00088	0.00431	0.00088
11	0.00782	0.00042	0.00543	0.00040

J	Mean Bias $d=-0.05$	MSE $d=-0.05$	Mean Bias $d=0.05$	MSE $d=0.05$
7	−0.02100	0.01152	−0.02322	0.01134
8	−0.00593	0.00439	−0.00864	0.00456
9	−0.00145	0.00185	−0.00447	0.00195
10	0.00169	0.00088	−0.00124	0.00090
11	0.00271	0.00039	−0.00032	0.00040

J	Mean Bias $d=0.15$	MSE $d=0.15$	Mean Bias $d=0.25$	MSE $d=0.25$
7	−0.02523	0.01135	−0.02817	0.01117
8	−0.01134	0.00486	−0.01481	0.00495
9	−0.00772	0.00209	−0.01121	0.00223
10	−0.00453	0.00094	−0.00819	0.00101
11	−0.00364	0.00042	−0.00723	0.00046

J	Mean Bias $d=0.35$	MSE $d=0.35$	Mean Bias $d=0.45$	MSE $d=0.45$
7	−0.03243	0.01104	−0.04605	0.00967
8	−0.01872	0.00501	−0.02701	0.00447
9	−0.01491	0.00237	−0.01962	0.00231
10	−0.01214	0.00112	−0.01627	0.00123
11	−0.01105	0.00054	−0.01497	0.00066

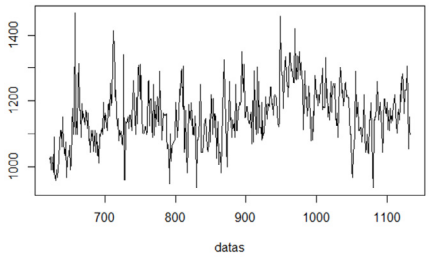
Table 2
Comparison of methods in observed time series.

Method	d	RMSE	MAPE	MAE
Nile River time series				
FrWT	0.45224	65.39579	0.04191	48.31246
GPH	0.42742	65.44793	0.04196	48.35560
Sperio	0.40890	65.52950	0.04204	48.44539
Whittle	0.38290	65.71103	0.04221	48.63099
CorrectRS	0.37565	65.77651	0.04228	48.70682
RoverS	0.24282	68.42912	0.04459	51.32852
Northern Hemisphere temperature time series				
FrWT	0.43391	0.57256	2.94470	0.41896
GPH	0.60732	0.58700	3.41043	0.42755
Sperio	0.48197	0.57433	3.06149	0.41968
Whittle	0.35870	0.57468	2.74745	0.42082
CorrectRS	0.45307	0.57302	2.99181	0.41910
RoverS	0.30576	0.58100	2.59334	0.42644
Ethernet time series				
FrWT	0.29171	1.27002	2.21990	0.86645
GPH	0.49101	1.30438	2.25542	0.81242
Sperio	0.44320	1.29104	2.24285	0.82180
Whittle	0.24597	1.27231	2.23009	0.88701
CorrectRS	0.33505	1.27208	2.22017	0.85030
RoverS	0.21549	1.27688	2.23895	0.90264

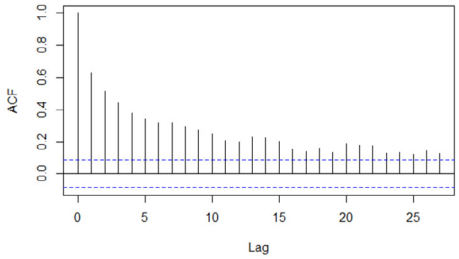
5. Empirical analyzes

In this section, we discuss the performance of the FrDWT-based estimator in observed time series. We also compare it with different approaches and heuristics to estimate the long memory parameter, as we did in the Simulations section.

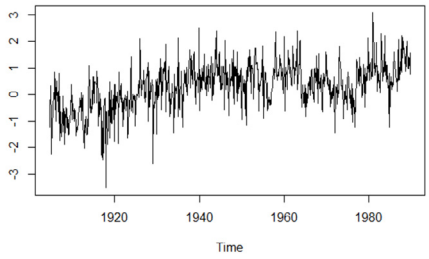
We consider three real time series that exhibit patterns of long range dependence. The first one is the yearly minimal water levels of the Nile River for the years 622 to 1133 at the Roda Gauge, near Cairo, made available in [8]. The second one is the monthly temperature series for the Northern Hemisphere for the years 1854-1989, as in [9]. The third one is the subset of the first



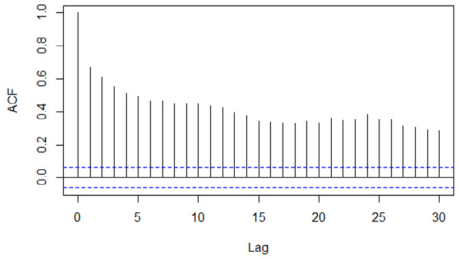
(a) Nile River yearly minimum level



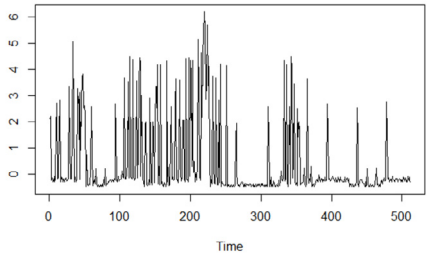
(b) Nile River yearly minimum level ACF



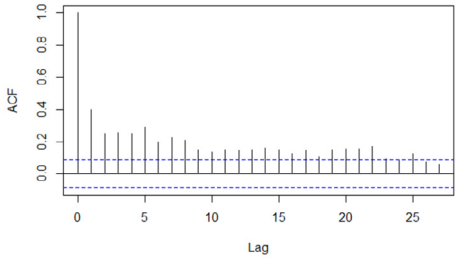
(c) Scaled Northern Hemisphere Temperature



(d) Northern Hemisphere Temperature ACF



(e) Scaled Ethernet traffic



(f) Scaled Ethernet traffic ACF

Fig. 9. Time series used in the empirical analyzes and their corresponding ACF.

512 observations of the Ethernet traffic time series from LAN at Bellcore, Morristown [46]. All of the time series were made available in [47]. We chose these series because, even though all of them have the traits of long-range dependent signals, they have diverse natures, since they com from different fields of knowledge: hydrology, climatology and technology.

Figs. 9a, 9c and 9e display the three analyzed series, whereas Figs. 9b, 9c and 9f show their respective ACF, side by side. It is possible to see the typical behavior of a long memory process from the slow, hyperbolic decay of their ACF.

For each of these time series, we fit a fractional Gaussian noise model and estimate the d parameter associated to it, following the FrDWT-based procedure. As means of comparison, we estimate the d parameter as well using the GPH estimator, the periodogram estimator (Sperio), the Whittle estimator, the rescaled range (R/S) estimator (RoverS) and the Corrected AL-R/S estimator (CorrectRS).

We compare these methods in terms of three metrics: RMSE, MAPE and MAE of the fitted series, which are defined as in [48]. With that, we aim to show that the proposed estimator not only performs well in simulated signals, but also in real, observed time series.

Table 2 exhibits the estimated value of d for each series and method and displays the corresponding RMSE, MAPE and MAE. We see that the FrDWT-based estimator outperforms every other method in terms of the three metrics for all series, except the GPH in terms of MAE in the Ethernet series and R/S in terms of MAPE in the Scaled Northern Hemisphere time series. This means that the FrDWT has the best performance in 7 out of 9 evaluations for the observed series, which provides a satisfactory performance and a competitive approach to analyze and model signals with long range dependence.

6. Conclusion and further remarks

We proposed a new estimator for the long memory parameter based on fractional spline wavelets based on extensions of Schoenberg's B-splines to fractional orders. The proposed estimator relies on the fact that, after performing the fractional spline wavelet transform on a pure long memory series, under a scaling factor passing through the wavelet channel, the series can be whitened.

The proposed estimator performed noticeably better in the simulated data than those we compared it with, namely: GPH, periodogram-based, Whittle, R/S, and AL-R/S estimators. In addition, the fractional spline wavelets based estimator had smaller variance and was less biased for all values of d. As for consistency, the estimator's bias and variance decrease as J increases in the simulated data, indicating good properties of consistency. Moreover, when comparing the MSE per sample size between estimators, we find that the FrDWT-based estimator outperforms the

other ones we compared it with for sample sizes greater than $T = 1024$ data points, whereas it is competitive with the Whittle estimator for small sample sizes ($T = 128$ and $T = 256$) and only for specific values of d. In general, in terms of MSE, the proposed estimator performs well.

When it comes to real, observed time series, the performance of the estimator is consistent with that presented in the simulation study, since it outperformed the other approaches in 7 out of 9 evaluations, in the three signals. Moreover, since the signals have different natures and originated from different fields of study, the results in the empirical analyzes section suggest that the proposed estimator can be used in various circumstances and a different range of applications as a competitive approach to the other methods that are widely used in the literature.

Fractional spline wavelets seem to be a promising tool to model data with fractal structure, analyze and synthesize long memory data, and pink noise whitening. We hope that other estimators or algorithms based on this whitening property may emerge from analyzing signals with this type of behavior.

Further work is necessary to study the asymptotic behavior of the estimator, deriving theoretical or numerical confidence intervals for hypothesis testing. It would also be interesting to assess the behavior of the proposed estimator when considering other types of data with a long memory and other distributions for the disturbance term. Moreover, due to the performance of the proposed estimator, it would be useful to develop an estimator for the ARFIMA(p,d,q) and for the FIGACH(m,D,n) models which are based on the FrDWT. These works will be carried elsewhere.

CRediT authorship contribution statement

Mateus Gonzalez de Freitas Pinto: Conceptualization, Methodology / Study design, Software, Formal analysis, Investigation, Resources, Data curation, Writing - original draft, Writing - review and editing, Visualization, Project administration. **Chang Chiann** Conceptualization, Methodology / Study design, Validation, Formal analysis, Resources, Writing - original draft, Writing - review and editing, Supervision, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Chang Chiann reports financial support was provided by State of São Paulo Research Foundation.

Data availability

Data will be made available on request.

Acknowledgments

C. Chiann would like to acknowledge the partial support of State of São Paulo Research Foundation - FAPESP through the grant 2018/04654-9.

References

[1] H.E. Hurst, Long-term storage capacity of reservoirs, Trans. Am. Soc. Civ. Eng. 116 (1) (1951) 770–799.
[2] B.B. Mandelbrot, J.W.V. Ness, Fractional Brownian motions, fractional noises and applications, SIAM Rev. 10 (4) (1968) 422–437.
[3] B.B. Mandelbrot, J.R. Wallis, Noah, Joseph, and operational hydrology, Water Resour. Res. 4 (5) (1968) 909–918.
[4] C.W.J. Granger, The typical spectral shape of an economic variable, Econometrica 34 (1) (1966) 150–161.
[5] J.R.M. Hosking, Fractional differencing, Biometrika 68 (1) (1981) 165–176.
[6] C.W.J. Granger, R. Joyeux, An introduction to long-memory time series models and fractional differencing, J. Time Ser. Anal. 1 (1) (1980) 15–29.
[7] J. Beran, Y. Feng, S. Ghosh, R. Kulik, Long Memory Process: Probabilistic Properties and Statistical Methods, Springer-Verlag, Germany, 2013.
[8] J. Beran, Statistics for Long-Memory Processes, Chapman and Hall, USA, 1994.
[9] P.D. Jones, K.R. Briffa, Global surface air temperature variations during the twentieth century: Part 1, spatial, temporal and seasonal details, Holocene 2 (2) (1992) 165–179.
[10] W. Palma, Long-Memory Time Series: Theory and Methods, Wiley and Sons Interscience, USA, 2007.
[11] E. Zivot, J. Wang, Modelling Financial Time Series with S-PLUS, 2nd edition, Springer, USA, 2005.
[12] R. Weron, Estimating long-range dependence: finite sample properties and confidence intervals, Phys. A, Stat. Mech. Appl. 312 (1) (2002) 285–299.
[13] A.A. Anis, E.H. Lloyd, The expected value of the adjusted rescaled Hurst range of independent normal summands, Biometrika 63 (1) (1976) 111–116.
[14] V.A. Reisen, Estimation of the fractional difference parameter in the arima(p, d, q) model using smoothed periodogram, J. Time Ser. Anal. 15 (3) (1994) 335–350.
[15] J. Geweke, S. Porter-Hudak, The estimation and application of long memory time series models, J. Time Ser. Anal. 4 (4) (1983) 221–238.
[16] H.-P. Graf, Long-range correlations and estimation of the self-similarity parameter, Ph.D. thesis, ETH, Zürich, 1983.
[17] M.J. Jensen, Using wavelets to obtain a consistent ordinary least squares estimator of the long-memory parameter, MPRA Paper 39152, University Library of Munich, Germany, 1999.
[18] M.J. Jensen, An alternative maximum likelihood estimator of long-memory processes using compactly supported wavelets, J. Econ. Dyn. Control 24 (3) (2000) 361–387.
[19] M. Unser, T. Blu, Fractional splines and wavelets, SIAM Rev. 42 (1) (2000) 43–67.
[20] M. Unser, T. Blu, Construction of fractional spline wavelet bases, in: M.A. Unser, A. Aldroubi, A.F. Laine (Eds.), Wavelet Applications in Signal and Image Processing VII, in: Proc. SPIE, vol. 3813, International Society for Optics and Photonics, 1999, pp. 422–431.
[21] I.J. Schoenberg, Contributions to the problem of approximation of equidistant data by analytic functions, Q. Appl. Math. 4 (1946) 45–99, 112–141.
[22] S.G. Mallat, Multiresolution approximations and wavelet orthonormal bases of l2(r), Trans. Am. Math. Soc. 315 (1) (1989) 69–87.
[23] T. Blu, M. Unser, The fractional spline wavelet transform: definition and implementation, in: 2000 IEEE International Conference on Acoustics, Speech, and Signal Processing, Proceedings (Cat. No. 00CH37100), vol. 1, 2000, pp. 512–515.
[24] S.G. Mallat, A Wavelet Tour of Signal Processing: The Sparse Way, Elsevier, USA, 2009.
[25] B. Vidakovic, Statistical Modeling by Wavelets, Wiley and Sons, USA, 1999.
[26] J. Ahlberg, E. Nilson, J. Walsh, The Theory of Splines and Their Applications, 1st edition, Mathematics in Science and Engineering, vol. 38, Academic Press, 1967.
[27] Y. Wang, Smoothing Splines: Methods and Applications (Chapman and Hall CRC Monographs on Statistics and Applied Probability), 1st edition, Monographs on Statistics and Applied Probability, vol. 121, Chapman and Hall/CRC, 2011.
[28] G. Micula, S. Micula, Handbook of Splines, 1st edition, Mathematics and Its Applications, vol. 462, Springer, Netherlands, 1999.
[29] J. Ramsay, G. Hooker, S. Graves, Functional Data Analysis with R and MATLAB, 1st edition, Use R, Springer-Verlag, New York, 2009.
[30] M. Unser, Splines: a perfect fit for signal and image processing, IEEE Signal Process. Mag. 16 (6) (1999) 22–38.
[31] M. Unser, T. Blu, Wavelet theory demystified, IEEE Trans. Signal Process. 51 (2) (2003) 470–483.
[32] C. Chui, J. Wang, On compactly supported spline wavelets and a duality principle, Trans. Am. Math. Soc. 330 (1992).
[33] C.D. Boor, A Practical Guide to Splines: With 32 Figures, rev. edition, Applied Mathematical Sciences, vol. 27, Springer, 2001.
[34] C.K. Chui, An Introduction to Wavelets, 1st edition, Wavelet Analysis and Its Applications, vol. 1, Academic Press, 1992.
[35] J. Liouville, Note sur une formule pour les différentielles à indices quelconques à 'loccasion d'un mémoire de M. Tortolini, J. Math. Pures Appl. 20 (unnumbered) (1855).
[36] Y. Meyer, Orthonormal wavelets, in: J.-M. Combes, A. Grossmann, P. Tchamitchian (Eds.), Wavelets, Springer Berlin Heidelberg, Berlin, Heidelberg, 1989, pp. 21–37.
[37] R.P. Brent, Algorithms for Minimization Without Derivatives, Prentice-Hall Series in Automatic Computation, Prentice-Hall, 1972.
[38] R.H. Shumway, D.S. Stoffer, Time Series Analysis and Its Applications with R Examples, fourth edition, Springer, USA, 2016.
[39] R.S. Tsay, Analysis of Financial Time Series, second edition, Wiley and Sons, USA, 2005.
[40] G.M. Ljung, Diagnostic testing of univariate time series models, Biometrika 73 (3) (1986) 725–730.
[41] G. Athanasopoulos, R.J. Hyndman, Forecasting: Principles and Practice, 2nd edition, Otexts, Australia, 2018.

[42] H. Hassani, M.R. Yeganegi, Selecting optimal lag order in ljung–box test, Phys. A, Stat. Mech. Appl. 541 (2020) 1–11.
[43] R Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2013, http://www.R-project.org/.
[44] J.Q. Veenstra, Persistence and antipersistence: theory and software, Ph.D. thesis, The University of Western Ontario, 2013.
[45] A.I. McLeod, H. Yu, Z.L. Krougly, Algorithms for linear time series analysis: with r package, J. Stat. Softw. 23 (5) (2007) 1–26.
[46] W. Leland, M. Taqqu, W. Willinger, D. Wilson, On the self-similar nature of ethernet traffic (extended version), IEEE/ACM Trans. Netw. 2 (1) (1994) 1–15.
[47] M. Maechler, longmemo: statistics for long-memory processes (Book Jan Beran), and related functionality, R package version 1.1-2, https://CRAN.R-project.org/package=longmemo, 2020.
[48] R. Hyndman, A. Koehler, Another look at measures of forecast accuracy, Int. J. Forecast. 22 (2006) 679–688.

Mateus Gonzalez de Freitas Pinto graduated with a bachelor's degree in Economics from the Federal University of ABC (2019) and received his M. Sc. in Statistics from the University of São Paulo (2021). He is currently pursuing his Ph.D. in Statistics in the University of São Paulo. His research interests include financial time series, long memory models, signal processing and wavelet analysis of time series. He is currently a risk analyst at the Brazilian Stock Exchange and OTC.

Chang Chiann graduated with a bachelor's degree in Statistics from the University of São Paulo (1989), master's degree in Statistics from the University of São Paulo (1993), Ph.D. from the University of São Paulo (1997) and a post-doctorate from the University of São Paulo (2000). She is currently an associate professor at the Institute of Mathematics and Statistics in the University of São Paulo. Her research interests include time series analysis, wavelets, analysis of bioavailability, bioequivalence studies, biosimilarity and clinical research.